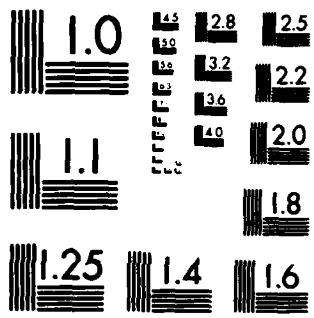
AD-A127 147 INTERACTIVE GRAPHICS PACKAGE TO FACILITATE A COMPARISON OF DATA WITH STAT..(U) ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND DOVER NJ COST A.. G E MAYERNIK 1/1

UNCLASSIFIED JAN 83 ARRAC-83-1 F/G 9/2 NL

END
DATE
FILMED
5 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

AD A127147

INTERACTIVE GRAPHICS PACKAGE TO FACILITATE

A COMPARISON OF DATA WITH STATISTICAL DISTRIBUTIONS

George E. Mayernik

Approved by:

ROBERT E. LEE
Chief, Cost Analysis Division
Requirements & Analysis Office

A

US ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
REQUIREMENTS AND ANALYSIS OFFICE
COST ANALYSIS DIVISION, HEADQUARTERS
DOVER, NEW JERSEY

January 1983

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A127 147
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) INTERACTIVE GRAPHICS PACKAGE TO FACILITATE A COMPARISON OF DATA WITH STATISTICAL DISTRIBUTIONS	5. TYPE OF REPORT & PERIOD COVERED Final
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) George E. Mayernik	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS ARRADCOM, Requirements & Analysis Office Cost Analysis Division (DRDAR-RAC) Dover, NJ 07801	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS ARRADCOM, TSD STINFO Division (DRDAR-TSS) Dover, NJ 07801	12. REPORT DATE
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Interactive computer program, graphics computer program, statistical distributions, probability distributions, histogram, frequency, bar charts, intervals, statistical density functions, probability density functions, uniform, normal, beta, weibull, rayleigh, exponential, parameter estimation, goodness-of-fit test, Kolomogorov-Smirnov test, one-sample test, chi-square test, Non-parametric stat tests

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the usage of an interactive FORTRAN computer program written for use on a Tekronix 4014 graphics terminal connected to a CDC 74/825/6600 computers under the NOS/BE operating system. The package is designed for visual analysis of data, testing the hypothesis that the data can be considered to be from one of several statistical distributions, density function parameter estimation, and performing one-sample goodness-of-fit tests. No prior computer experience is necessary to run this program; however, some statistical knowledge is required to interpret the output. A source code listing of the program may

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. be obtained from the performing organization. Program and/or job control language modifications may be necessary for other than CDC computers and/or NOS/BE operating systems.

DISCLAIMER

A

TABLE OF CONTENTS

TABLES

FIGURES

INTRODUCTION

This report was prepared to assist scientists and engineers in using the interactive graphics statistical data analysis computer program package.

The package was written to: analyze data visually via relative frequency histogram(s) and/or superimposed probability density function(s), test the hypothesis that the data is from one of six well known continuous statistical distributions, estimate density function parameter(s) by method of maximum likelihood or method of moments, and perform chi-square and/or Kolmogorov-Smirnov one-sample goodness of fit test(s).

The program provides the user with six major options and several other options which are contained within the major options. The options are flexible and easy to use. The user selects:

1. The number of histogram intervals,

2. User or computer scaling of histogram,

3. Statistical program control,

4. Which statistical distribution to fit to the data,

5. Whether the user inputs the distribution parameters or the computer estimates the parameters,

6. Whether or not to perform goodness-of-fit test(s).

A seperate interactive FORTRAN data file creation program which reformats the data to be used as input to the interactive graphics package is also described in this report. This program was written for use on CDC 74/825/6600 computers under the NOS/BE operating system.

No prior computer experience is necessary to run either program. Details of both computer programs will be discussed in their respective write up and examples of each program are provided.

Any comments for improvements or corrections to these programs, requests for source code listings, or requests for additional statistical distributions to be added should be directed to Commander, USA Armament Research and Development Command, Requirements and Analysis Office, Cost Analysis Division (DRDAR-RAC), Dover, NJ 07801.

DATA FILE CREATION PROGRAM

This interactive computer program prompts the user to enter various input data, reformats the data, and writes it on a computer file. After terminating the program, the user stores the data file into the computer to be used by the interactive graphics statistical data analysis package. Since inputting large data files can be time consuming, no graphics capability is necessary to create data files, and there are fewer Tekronix 4014 graphics terminals than non-graphical electronic data terminals, the data file creation program was not included as part of the graphics statistical package.

You will need a valid user name and password to use any of the computer facilities. Contact the Management Information Systems Directorate (MISD) X4714 for further instructions on how to get a user name and password.

Plug in and turn on the terminal. If you terminal requires use of a telephone, contact a member of MISD, Techniques Division for a list of current ARRADCOM computer telephone numbers. Dial one of the numbers, listen for a high pitch tone, then place phone in holder noting the direction to place the mouthpiece as indicated by the directions or picture on the terminal, to access the computer. If the terminal has a Gandolf modem accompanied with it, set the dial on the Gandolf modem to 55 or 75 (2400 baud rate), 54 or 74 (1200 baud rate), or 53 or 73 (300 baud rate); make sure the terminal switches are set for the appropriate baud rate; and raise the switch to the up position to access the computer.

The following general instructions will help you run the program quickly and with minimal effort:

- Remember to hit the "RETURN" key after each line entry.

- Most of the numerical program inputs are in free format, however, when asked to input an integer value do not use a decimal point or an error message will be written.

- Leading and trailing zeroes do not have to be entered.

- To erase[1] the last character typed, simultaneously press the "CONTROL" and "H" keys. Then retype the correct character.

[1] The last character is not physically erased. The carriage moves back one space and the user types over the previously entered character.

● To erase[2] an entire line (before you hit the carriage return) simultaneously press the "CONTROL" and "X" keys. Retype the entire corrected line.

● You will be asked if you wish to make any corrections at the completion of the input phase.

● When the computer run is complete, it will type "COMMAND".

Appendix A shows the user how to: access the data file creation program, input the data, display the data, modify the data, terminate the program, store the data file, and get off the computer. All user inputs have been circled for demonstrative purposes. The user is allowed a maximum of 30 characters for the horizontal axis label and diagram title. The vertical axis label has a restriction of 15 characters or less.

If the user wishes to analyze more then one set of data without having to reexecute the interactive graphics statistics package program, two or more data sets may be placed on one data file. This can be accomplished by reexecuting the data file creation program after terminating the program the first time and before cataloging the data file. The statisics package program takes a longer time to execute than the data file creation program, thus sometimes the user may want to store several data sets on one data file. This procedure is shown in Appendix B.

---

2 The line is not physically erased off the paper, however, it is erased from the buffer memory.

INTERATIVE GRAPHICS STATISTICS PACKAGE

In many fields, certain random variables are known from past history to have a particular statistical distribution. However, often the engineer or scientist does not know and/or falsely assumes the distribution of a random variable. This computer program provides a visual as well as numerical aid for comparing data to common probability density functions. This additional information will help guide the user to select the best statistical distribution for the given random variable. Reference 1 is excellent for defining statistical terminology used in this report.

The following are examples of the program's usefulness. An engineer in a laboratory may have miss distance data from test firings and would like to know whether the data can be represented by a common statistical distribution. In a second example, a cost analyst may wish to develop a Cost Estimating Relationship (CER) which is an equation that related costs of an item to one or more physical or performance characteristics. CER's are for the most part developed through applications of regrssion analysis. Usually in regression analysis, T-tests and F-Tests are performed which require the assumption that the residuals follow a normal distribution. Using the statistics package program can help determine whether the normality assumption is satisfied. An outstanding text on regression analysis is reference 2.

After using the data file creation program and cataloging the data file, the interactive graphics statistics package can be used. The package was written to: analyze data visually by drawing relative frequency histogram(s) and/or superimpose probability density functions on the histogram(s), test the hypothesis that the data is from one of six common continuous statistical distributions via chi-square and/or Kolomogorov-Smirnov one-sample goodness of fit test(s), and estimate desnity function parameters using the data by the method of maximum likelihood or the method of moments.

The following general instructions may help you run the program quickly with minimal effort:

- A bell will prompt the user each time user action is required.

- When asked to input numerical values, hit the "RETURN" key after the entry to continue execution of the program.

- When asked to select an option with the cursor, hit the blank key or any character key after selecting an option with the horizontal cursor to continue execution of the program.

- Each time an option is selected with the cursor, an arrow appears pointing to the selection.

- All of the numerical program inputs are in free format.

- Leading or trailing zeroes do not have to be entered.

- To erase[1] the last character typed, simultaneously press the "CONTROL" and "H" keys. Then retype the correct character.

- To erase[2] an entire line (before you hit the carriage return) simultaneously press the "CONTROL" and "X" keys. Retype the entire corrected line.

- If a Hard Copy Unit is accompanied with the terminal, the copy switch may be used at any time during the session to obtain permanent high contrast copy of the terminal display image.

- If a stop error message occurs while running the program, see table 1 to correct the error.

We will now go through a step-by-step example to see how the program works. All features of the program will not be discussed in this example. The remainder of the features will be commented on later in the report.

Figure 1 shows the user how to access the interactive graphics statistics package program. User inputs are circled for clarity.

After typing GRSTAT and hitting the carriage return, the screen will flash and pause several times, then figure 2, requesting the user to input the number of intervals for the histogram, will appear on the screen. Simply enter a number from 2 to 80.

Next the user is asked to select with the horizontal cursor whether or not to allow the computer to scale the X axis of the histogram (see figure 3). Using the thumb wheel located on the right side of the keyboard, the horizontal cursor was moved over the words "COMPUTER SCALING FOR X AXIS" and the blank key or any other character is hit to continue program execution.

At this point, the histogram is drawn and the statistics program control option is displayed as in figure 4. If the user has put several data sets on one file, as discussed in the data file creation program section, the user may now analyze the next set of data. A second option is to modify the current histogram. A third option is to fit one of six well known statistical distributions to the histogram data. Finally, the user may terminate the program. In this example, "NEW DENSITY FUNCTION" option was selected.

---

[1] Ibid.
[2] Ibid.

Figure 5 appears on the screen requesting the user to select among the six statistical distributions. Information about each of the six statistical distributions is given in table 2 and will be discussed later on in the report. The normal distribution was selected.

Figure 6 is displayed giving the user the option of inputting the distribution parameters. "COMPUTER ESTIMATES PARAMETERS" option was selected. The method of maximum likelihood was used to derive all parameter estimators with the exception of the beta distribution parameters which were derived using the method of moments. The estimation formulas are in table 2. Thus, using the given data, the computer estimates the values of mean and standard deviation for the normal distribution by the method of maximum likelihood.

The probability density function, parameter space, parameter values, and superimposed normal probability density function are written on the screen. Then, the user may choose to perform a Kolmogorov-Smirnov one-sample test or a chi-square goodness of fit test to test the hypothesis that the data is from, in this case, a normal distribution (see figure 7). A chi-square test can not be performed with less than 10 data points, since this will cause the number of degrees of freedom in the chi-square comparison to be less than one. Also, if one parameter is estimated, 15 data points are needed and if two parameters are estimated, 20 data points are needed, otherwise, the number of degrees of freedom will be less than one. In this example, the chi-square goodness of fit test was selected. One speaks of "goodness of fit" when one tries to compare an observed frequency distribution with the values of a theoretical distribution. If the "COMPUTER ESTIMATES PARAMETERS" option was selected, the user is asked to enter the theoretical distribution parameters for the continuous distribution previously selected. In most cases, the user will want to enter the values used to compute the probability density function superimposed on the histogram; i.e., those values printed in the lower left hand corner of the screen (see figure 8). If the "USER INPUTS PARAMETERS" option was selected, the computer uses the inputed values as the theoretical distribution parameters.

The computer calculates and prints the chi-square test result in figure 9. The value printed is the probability of the chi-square table value exceeding the computed chi-square statistic, if the null hypothesis of equality of the actual distribution of the observed data and the selected continuous distribution is true (against the alternative of inequality of the distributions); i.e., the value printed is the probability of accepting the null hypothesis that the data is from the selected distribution (against the alterntive that the data is from some other distribution), if the data is from the selected distribution. Since the value 0.5918 exceeds typical significance levels (0.01, 0.05, etc.), one usually would accept the null hypothesis that the data is normally distributed. The interpretation of Kolmogorov-Smirnov test results are discussed on page 7 of the report. More information about both tests may be obtained in reference 3.

After the chi-square test result is displayed, the user is once again asked to select among a Kolomogorov-Smirnov test, chi-square test, or not to perform any test. If "NO TEST" is selected, an arrow is pointed to "NO TEST" and the user is returned to the statistics program control option (see figure 10). After selecting "TERMINATE THE PROGRAM" and typing "LOGOUT", figure 11 is displayed.

A few of the options previously mentioned will now be discussed in more detail.

The user scaling of histogram option allows the user to adjust the scaling to his or her needs or liking. When used in conjuction with the number of intervals for histogram option, the user can get a histogram with even increments for report purposes, which is sometimes not obtained using the computer scaling option. For example, this feature can be used to make certain Weibull, Rayleigh, and exponentially distributed data have a minimum histogram scale value of zero. After selecting this option, the user is asked to enter the lower (minimal value -0.001) and upper (maximum value 10^5) limits of the histogram as can be seen in figure 12.

The user inputted distribution parameters option becomes useful for those distributions that have a critical value of a parameter which may produce a sharp change in the shape of the curve around that particular parameter value. After selecting this option, the user is requested to enter values for each parameter of the continueous distribution selected. This can be seen in the lower left hand corner of figure 13.

The Kolmogorov-Smirnov one-sample test option, like the chi-square goodness of fit test option, helps the user determine whether the sample data fits the theoretical distribution previously selected. The Kolomogorov-Smirnov (K-S) one-sample test is generally more efficient than the chi-square test for goodness of fit for small samples, and it can be used for very small samples where the chi-square test does not apply. The K-S test is a nonparametric test for differences between cumulative distributions. This test concerns the agreement between an observed cumulative distribution of sample values and a specified continuous distribution function; thus, it is a test of goodness of fit. The program's method of calculations for the K-S test option is different depending upon whether the number of data points is ≤ 80 or >80. More accurate test results are computed and the test result is interpreted simular to the chi-square test result when the number of data points is >80. The value printed is the probability of the table value exceeding the computed test statistic, when the null hypothesis of equality of the actual distribution of the observed data and the selected continuous distribution is true (against the alternative of inequality of the distributions); i.e., the value printed is the probability of accepting the null hypothesis that the data is from the selected distribution (against the alternative that the data is from some other distribution), when the data is from the selected distribution. A printed value of 0.9996 exceeds typical significance levels (0.01, 0.05, etc.), thus one normally would accept the null hypothesis that the data is from the selected distribution. If the number of data points is $\leq$ 80, the user is asked to enter a K-S significance level (which must be .20, .15, .10, .05, or .01) to be used by the program to look up critical values for the K-S one-sample statistic. The null hypothesis of equality of the distributions (against a two-sided altnerative) will be rejected at the level of significance α, if the test statistic exceeds the $1-\alpha$ quantile of the K-S table. The computer will print out the test result, either accepting or rejecting the null hypothesis at the significance level selected by the user (see figure 14). Thus, the null hypothesis that the DIVAD miss distance data is normally distributed, is accepted at level 0.20. The user should attempt to get the test accepted at the highest level possible.

Each of the six common statistical distributions will now be commented on in more detail. A summary of the six parametric families of distributions can be found in table 2.

The uniform distribution gets its name from the fact that its density is uniform, or constant, over the interval [a,b]. It is also called the rectangular distribution - the shape of the density function is rectangular. When one speaks of a random number from an interval [0,1], one is thinking of the value of a uniformly distributed random variable over the interval [0,1]. The general shape of the distribution can be seen in figure 15. Many times the endpoints of the density function lie within histogram cells; however, the smallest and largest points in each of these cells are within the minimum and maximum points of the density function. The user may input distribution parameters which caused the endpoints of the density function to go off the histogram scale. This may be corrected by using the user scaling of histogram option.

A great many of the techniques used in applied statistics are based upon the normal distribution. It is also the limiting form of many probability distributions. "Grading on the bell shaped curve" is a classic example of using the normal distribution. Many times engineers and scientists falsely assume a random variable to be normally distributed. This program can aid technical people to decide whether their data is normally distributed. The general shape of the normal density can be seen in figure 16. Sometime the user may wish to display the asymtotic portions of this distribution on the histogram drawing. This can be accomplished by utilizing the user scaling of histogram option. Since the histogram is drawn before the distribution is superimposed, it is possible that the normal density may extend off the graph in the vertical direction and consequently cannot be plotted on the graph using the number of intervals chosen. When this occurs, the user is given the option of redefining the number of intervals or allowing the computer to rescale the graph.

The family of beta densities is a two-parameter family of densities that is positive on the interval (0,1) and can assume quite a variety of different shapes (see figure 17). Consequently, the beta distribution can be used to model an experiment for which one of the shapes is appropriate. An example of using the beta distribution arises in estimating the probability p, known as reliability that a missile will successfully fulfill a given mission. The initial information (before testing) concerning reliability based upon past experience, component data, engineering judgement, etc., might lead one to select a particular member of the beta family as the best a priori description of reliability. Information from actual missions is then added to the a priori model to obtain a revised "posterior" distribution, which is also a beta distribution, but with revised parameters. The number of intervals for histogram option and the user scaling of histogram option allow the user to get even increments beginning at zero and ending at one. Inputting negative parameter values when using the user inputs parameters option will cause the program to terminate, since the parameter space for the beta distribution is positive.

The Weibull distribution has been successfully used in reliability theory. The time-to-failure of many electronic and metal components have Weibull distributions. The Weibull density function with parameter C=2 is known as the Rayleigh density function. A plot of the Weibull density using the fictitious DIVAD miss distance data, as seen in figure 18, may be closer to a Rayleigh distribution, since the estimated value of the C parameter is near two. For B=1, the Weibull density reduces to the exponential density. Negative parameter values are not allowed when using the user inputs parameters option, since the parameter space is positive.

The Rayleigh distribution represents radial error when the error on two mutually perpendicular axes is normally distributed with equal variances around zero. This might arise in bombing problems, when the sighting errors in the X and Y directions are independent and we desire to evaluate the distribution of distance of the impact from the target. A second application arises in statistical communication theory. When random noise is detected by a linear dector the amplitude of the envelope of noise is distributed according to the Rayleigh distribution. Once again, the user must input a parameter value of B greater than zero. Figure 19 contains a plot of the Rayleigh density using the fictitious DIVAD test data.

The (negative) exponential distribution has been used as a model for lifetimes of various things. If the lifetime of a component has an exponential distribution, the conditional probability that the component will last A+B time units given that it has lasted A time units is the same as its initial probability of lasting B time units. Another way of saying this is to say that an "old" functioning component has the same lifetime distribution as a "new" functioning component or that the component is not subject to fatigue or to wear. It can be shown that in connection with Poisson process the waiting time between successive arrivals (successes) have exponential distributions. It has been shown theoretically that this distribution provides a reasonable model for systems designed with a limited degree of redundancy and made up of many components none of which has a high probability of failure, especially when low component failure rates are maintained by periodic inspection and replacement, or in situations where failure is a function of outside phenomena rather than the amount of previous stress. On the other hand, the exponential distribution often can not represent individual component life (because of "infant mortalities" and wear-out patterns), and it is sometimes questionable even as a systems model. Positive values for the parameter B must be inputted. A plot of an exponential density can be seen in figure 20.

A good method to use when the user has no idea which density the data may come from, is to fit each possible statistics package program distribution with the computer estimates parameter option and both goodness of fit tests. Then, select the best density function based on the two tests. Figures 15, 16, 18, 19 and 20 contain histograms, density function sketches, parameter estimates, and both chi-square and Kolmogorov-Smirnov goodness of fit test results for each possible statistics package program distributions that can be fit to the fictitious DIVAD miss distance data. Looking at the five superimposed density functions, it appears as though the normal, Weibull, and Rayleigh distributions fit quite well. The K-S test results also favor these three densities, since they all accept the null hypothesis at the highest possible level (0.2). The chi-square test leads the user to select the Rayleigh distribution because of the 0.9770 acceptance level. Thus, it is highly likely that the ficticious DIVAD miss distance data is from a Rayleigh distribution with B approximately equal to 2.5 as its parameter value.

REFERENCES

1. Mood, A., Gaybill, F., and Boes, D., Introduction to the Theory of Statistics, Third Edition, McGraw-Hill, Inc., 1974.

2. Draper, N.R. and Smith, H., Applied Regression Analysis, Second Edition, John Wiley and Sons, Inc., 1981.

3. Miller, I. and Freund, J., Probability and Statistics for Engineers, Second Edition, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1977.

Table 1. Correcting stop error messages

Stop error messages	Why error occured	How to correct error
STOP 11	Computer scaling option was selected. The width of the historgram interval is greater than or equal to 10^{10}. Program can not handle this extreme data without modification.	Omit extremely large and/or small data values (outliers) from the analysis; or increase number of intervals for histogram.
STOP 12	Computer scaling option was selected. The width of the historgram interval is less than 10^{10}. Program can not handle this extreme data without modification.	Decrease the number of intervals for histogram.
STOP 13	User did not create and/or attach proper data file prior to running statistics package program; or user selected "NEXT SET OF DATA" option and there was no more data on the file.	Use data file creation program and/or attach proper data file prior to running statistics package program; or add more data to the file.
STOP 20	Kolomogorov-Smirnov (K-S) one sample test was selected and number of data points was less then or equal to 80. Thus, user was required to input a K-S significance level and entered a value which is not allowed.	Rerun the program and enter one of the following K-S significance levels: .01, .05, .10, .15, or .20.
STOP 24	The program attempted to perform a chi-square (C-S) goodness of fit test with an expected cell frequency less than five. There is an insufficient amount of data to perform the C-S test.	Add more data to the analysis or do not perform the C-S test.

Table 2. Summary of parametric families of distributions

Name of parametric family of distributions	Probability sensity function $f(\cdot)$[a]	Parameter space	Mean[b]	Variance[b]	Method of finding estimator	Estimator[c]
Uniform or Rectangular	$f(X) = \frac{1}{b-a} I_{[a,b]}(X)$	$-\infty<a<b<\infty$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	maximum likelihood	$\hat{a}$ = minimum x value $\hat{b}$ = maximum x value
Normal	$f(X) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-(X-\mu)^2/2\sigma^2\right]$	$-\infty<\mu<\infty$ $\sigma<0$	μ	σ^2	maximum likelihood	$\hat{\mu} = \bar{x}, \hat{\sigma} = (1/n)\sum_{i=1}^{n}(x_i-\bar{x})^2$
Beta	$f(X) = \frac{1}{B(a,b)} X^{a-1} (1-X)^{b-1} I_{(0,1)}(X)$	$a > 0$ $b > 0$	$\frac{a}{a+b}$	$\frac{ab}{(a+b+1)(a+b)^2}$	moments	$\hat{a} = \left[\bar{x}^2(1-\bar{x})/S^2\right] - \bar{x}$ $\hat{b} = \left[(\bar{x}-2\bar{x}^2+\bar{x}^3)/S^2\right] + \bar{x} - 1$
Weibull	$f(X) = (cX^{c-1}/b^c) \exp\left[-(X/b)^c\right] I_{(0,\infty)}(X)$	$b > 0$ $c > 0$	$b\Gamma[(c+1)/c]$	$b^2\left(\Gamma[(c+2)/c] - \{\Gamma[(c+1)/c]\}^2\right)$	maximum likelihood	$\hat{b} = \left[(1/n)\sum_{i=1}^{n} x_i^{\hat{c}}\right]^{1/\hat{c}}$ $\hat{c} = n / \left[(1/\hat{b})^{\hat{c}} \sum_{i=1}^{n} x_i^{\hat{c}} \log_e x_i - \sum_{i=1}^{n} \log_e x_i\right]$
Rayleigh	$f(X) = (2X/b^2) \exp\left[-(X/b)^2\right] I_{(0,\infty)}(X)$	$b > 0$	$b\frac{\sqrt{\pi}}{2}$	$b^2(1-\frac{\pi}{4})$	maximum likelihood	$\hat{b} = \left[(1/n)\sum_{i=1}^{n} x_i^2\right]^{1/2}$
Exponential	$f(X) = (1/b) \exp\left[-(X/b)\right] I_{(0,\infty)}(X)$	$b > 0$	b	b^2	maximum likelihood	$\hat{b} = \bar{x}$

[a] $B(a,b) = \int_0^1 X^{a-1}(1-X)^{b-1}\,dX$ for $a>0$, $b>0$ is called the beta function.

[b] $\Gamma(t) = \int_0^\infty X^{t-1} e^{-X}\,dX$ for $t>0$ is called the gamma function.

[c] $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$ and $S^2 = (1/[n-1])\sum_{i=1}^{n}(x_i - \bar{x})^2$

```
CONTROL DATA INTERCOM 5.1
DATE  08/02/82
TIME  10.44.01.

PLEASE LOGIN                                   YOUR OWN USER NAME
LOGIN,LTEMYD3300,SUP
[redacted] ENTER PASSWORD-                     YOUR OWN PASSWORD
ENTER HARDWARE ID-LKA                          HARDWARE ID OF THE TERMINAL YOU ARE USING

SUP       CREATED-07/27/82   TODAY IS 08/02/82
  RELOAD OF PERMANENT FILES

    TYPE SYSBULL RELOAD FOR DETAILS

COMMAND- ETL,500
COMMAND- ATTACH,GRSTAT,TEKSTAT30,ID=DRDAR,MR=1
 AT CY= 001 SN=PFSET
COMMAND- ATTACH,TAPE10,KSTABLE,ID=DRDAR,MR=1
 AT CY= 001 SN=PFSET
COMMAND- ATTACH,TAPE5,DIVADDATA,ID=MAYERNIK,CY=1
 AT CY= 001 SN=PFSET                           CYCLE NUMBER OF THE DATA FILE
COMMAND- GRSTAT                                OWNER OR CREATOR OF FILE
                                               PERMANENT FILE NAME OF THE DATA FILE
```

Figure 1. Accessing the statistics program

NUMBER OF INTERVALS FOR HISTOGRAM = 10

Figure 2. Inputting number of histogram intervals

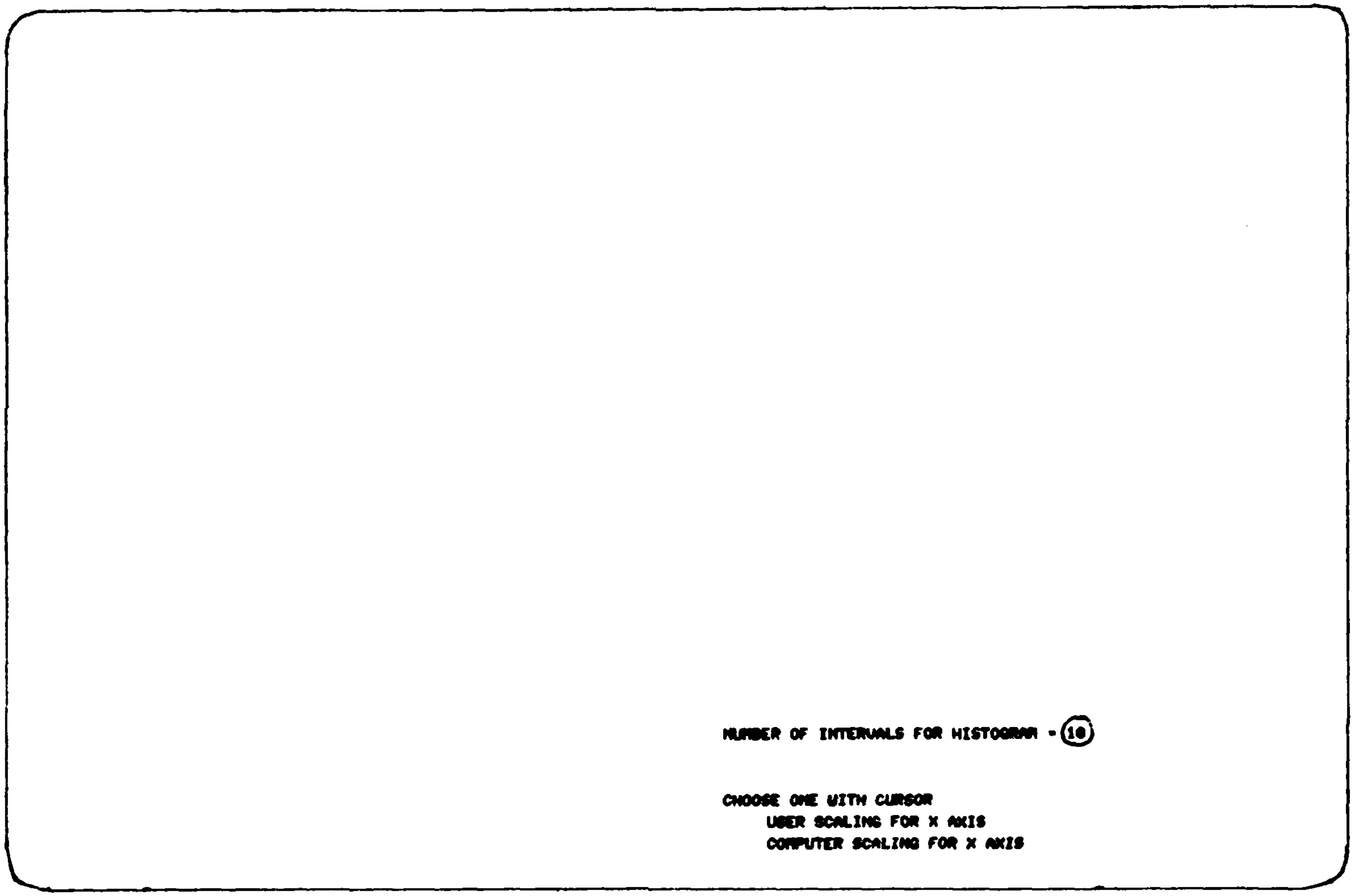


Figure 3. Histogram scaling option

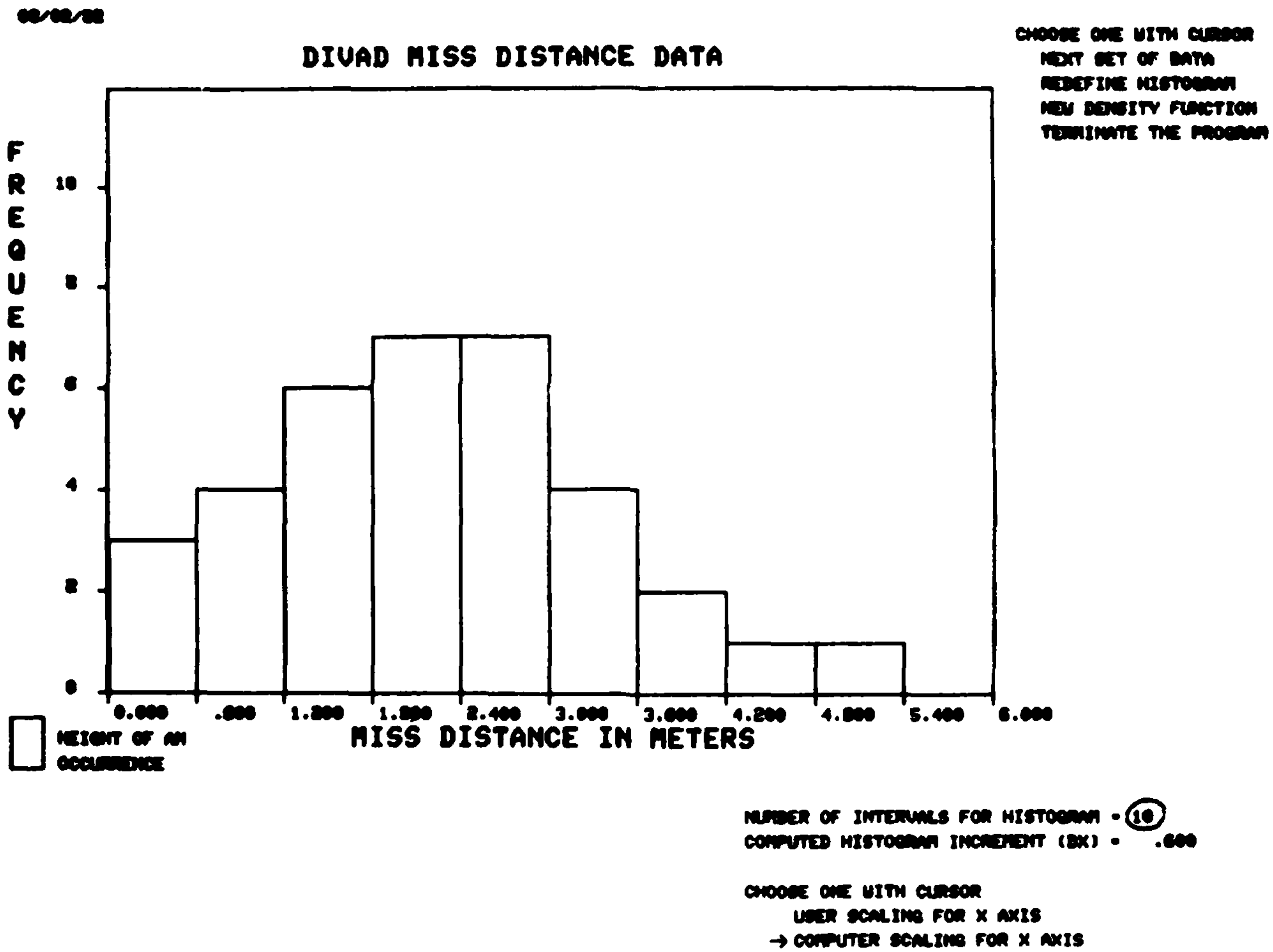


Figure 4. Statistics program control option

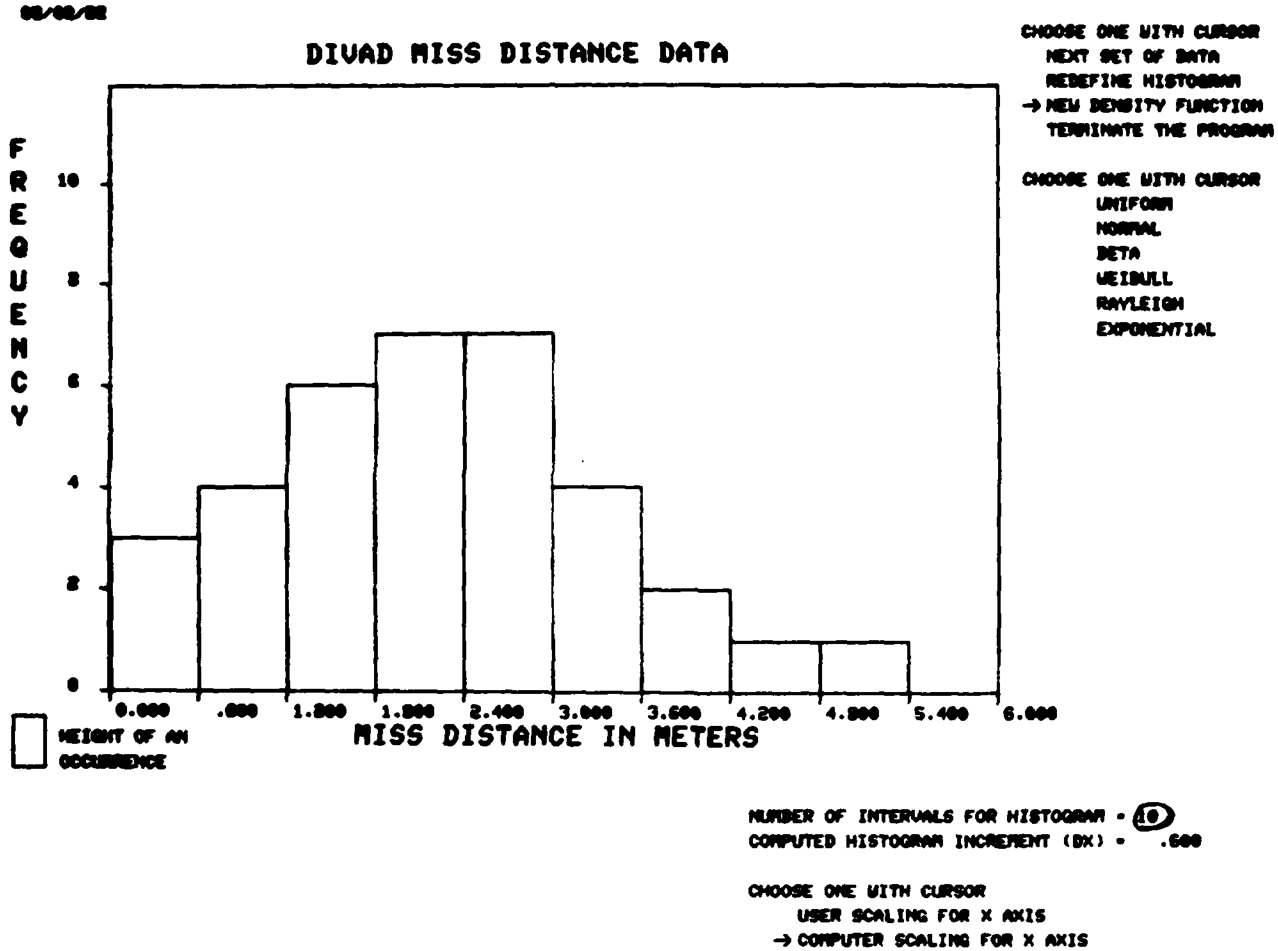


Figure 5. Statistical distribution option

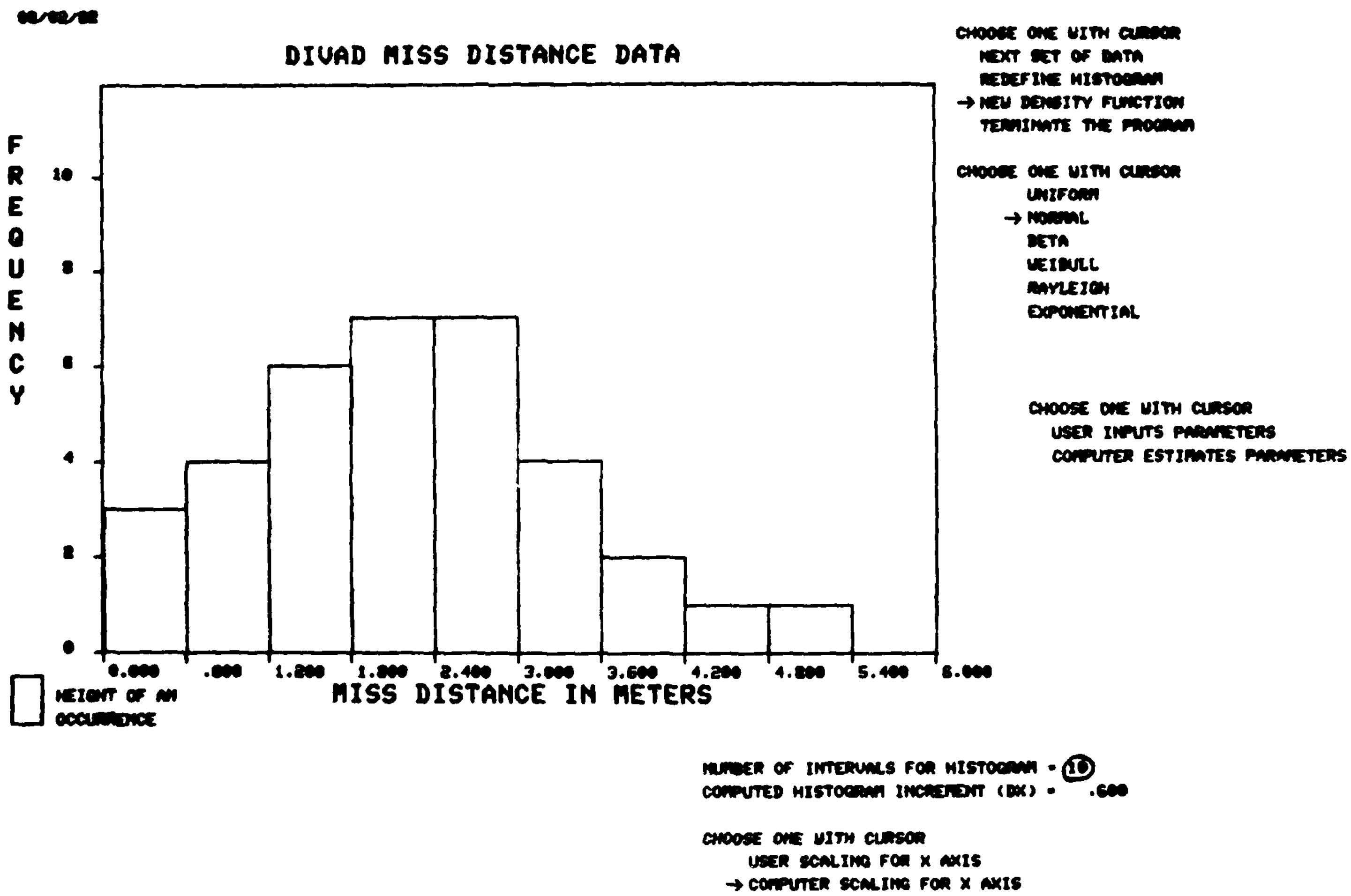


Figure 6. Statistical distribution parameter estimating option

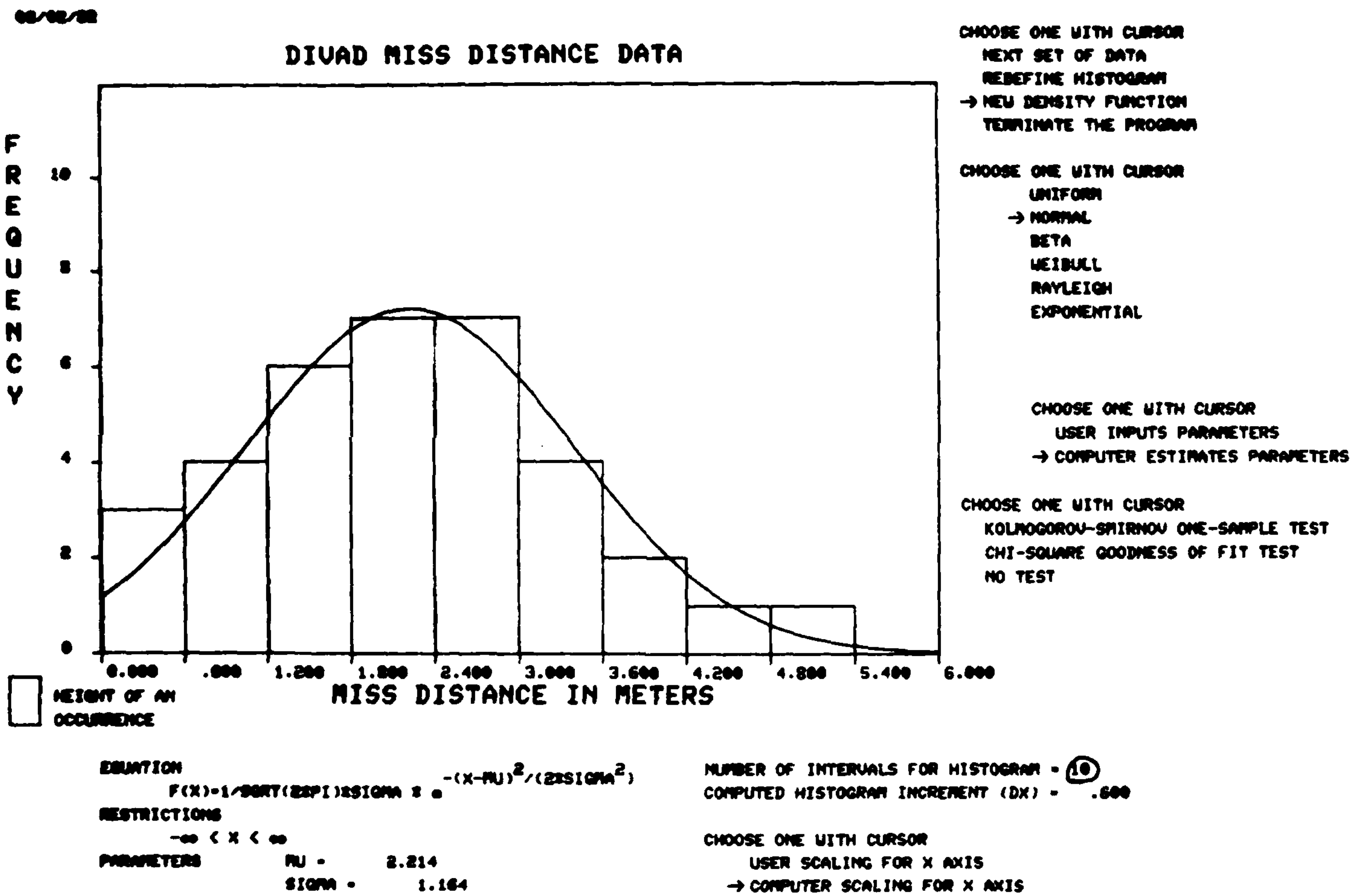


Figure 7. Goodness-of-fit test option

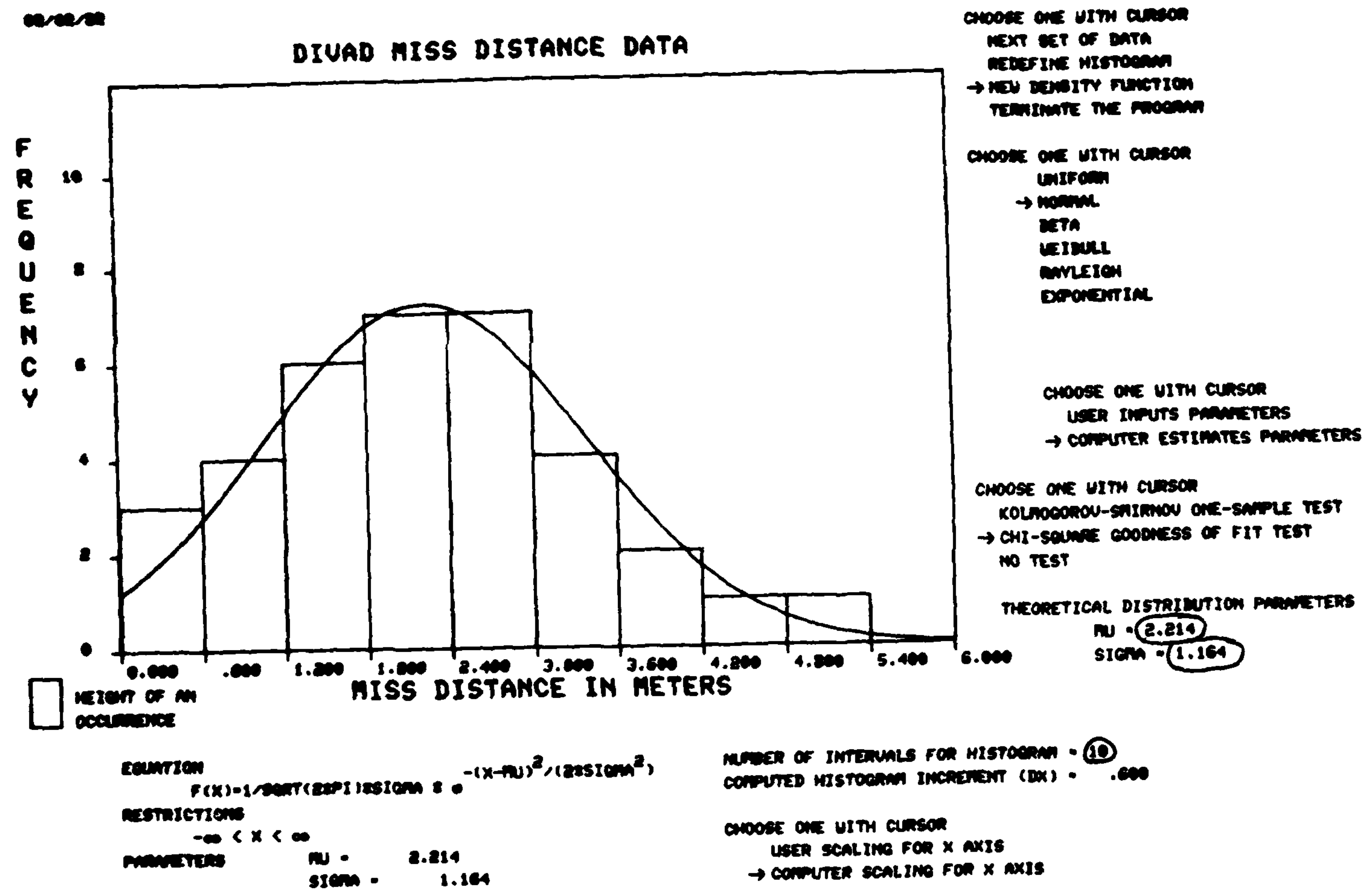


Figure 8. Inputting theoretical distribution parameters

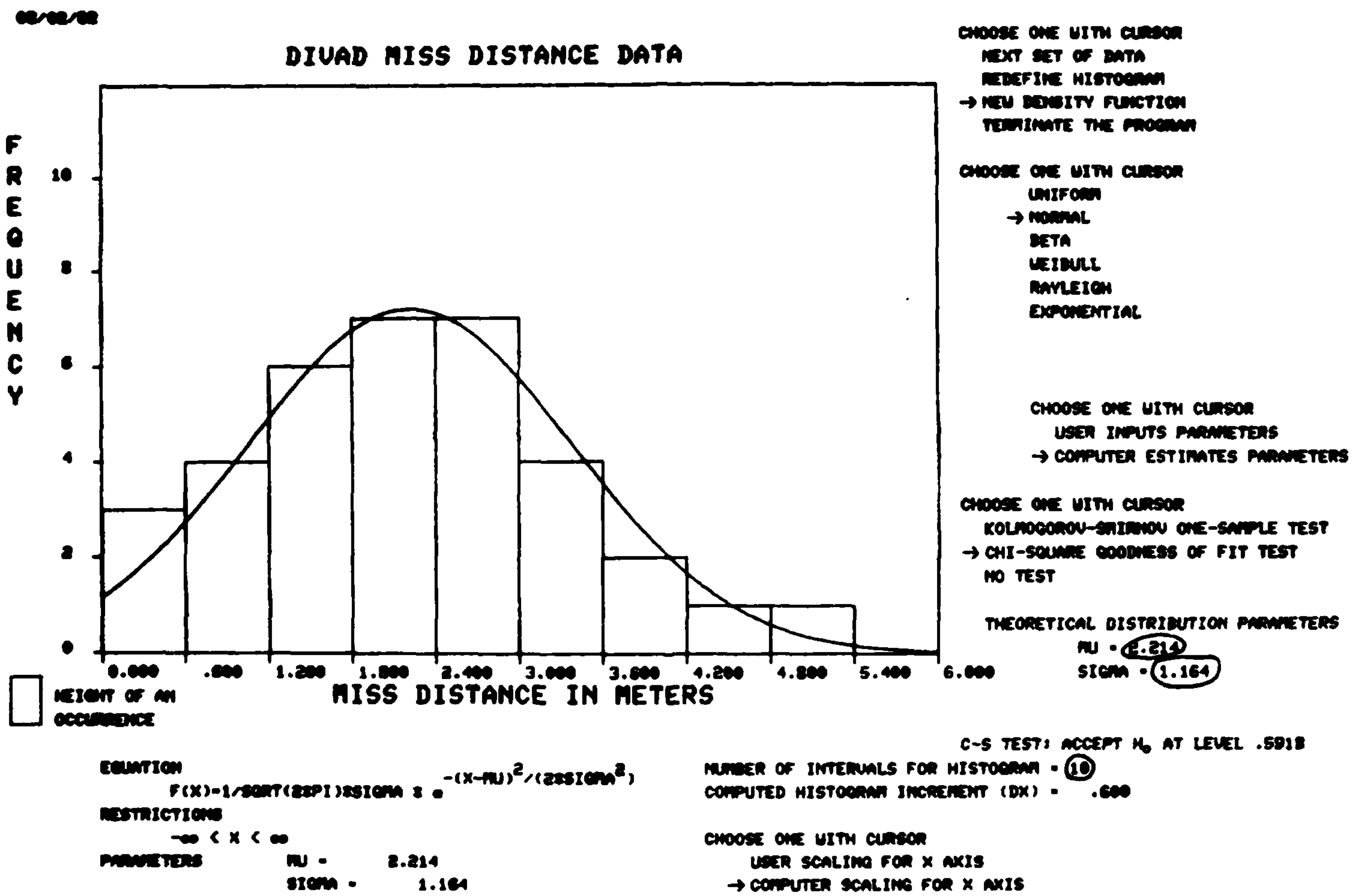


Figure 9. Chi-square test result

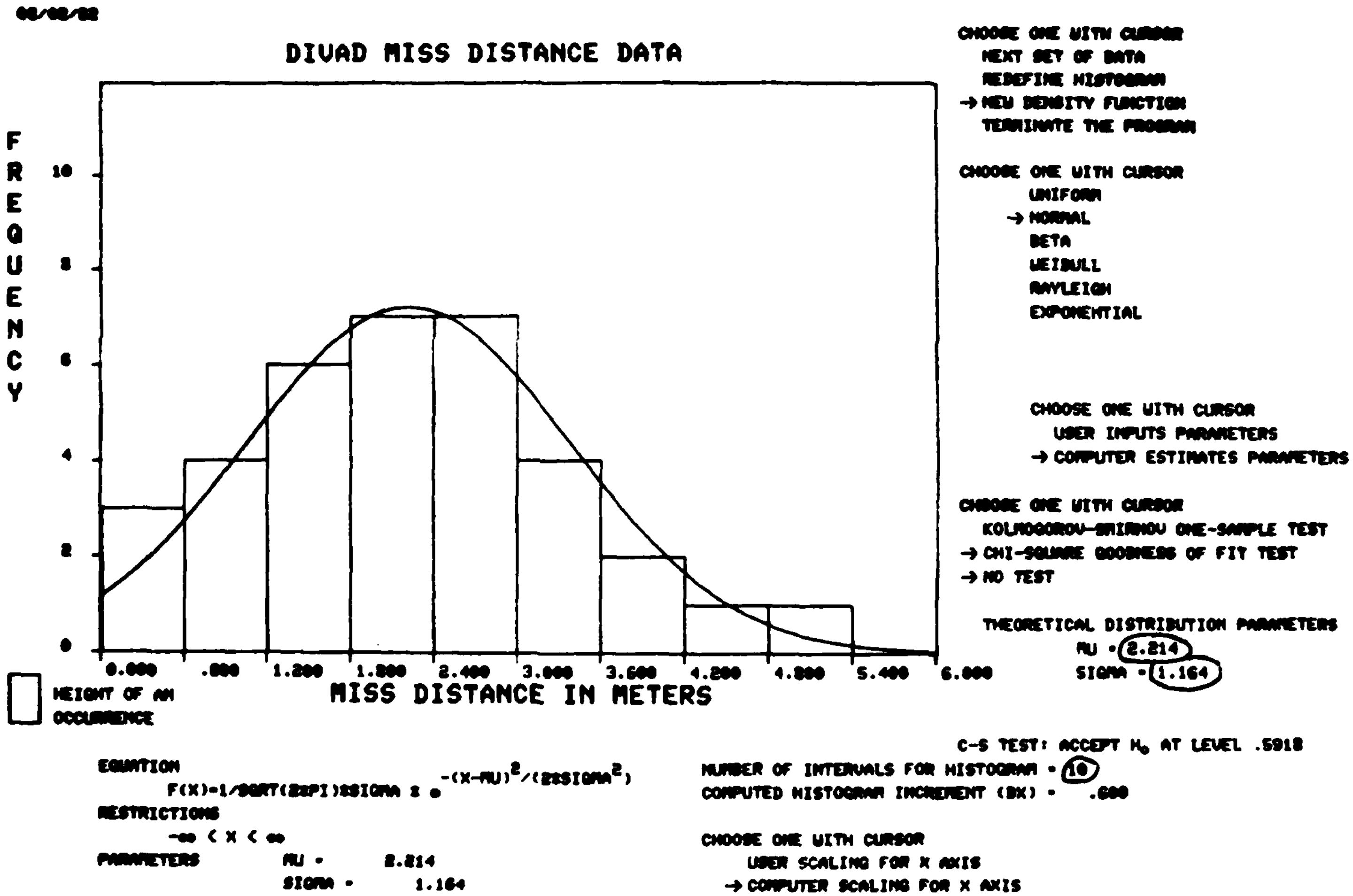


Figure 10. Returning to statistics program control option

```
     STOP  - PROGRAM HAS BEEN TERMINATED
     044100 MAXIMUM EXECUTION FL.
      2.100 CP SECONDS EXECUTION TIME.
COMMAND- LOGOUT
CPA         [illegible] SEC.         2.980 ABJ.
CPB          .000 SEC.          .000 ABJ.
SYS TIME                       5.172
EST. COST AT [illegible]/HR. - $       .35
CONNECT TIME     0 HRS.   23 MIN.
 [illegible]  LOGGED OUT AT 11.08.08.
      <
```

Figure 11. Terminating the program and getting off the computer

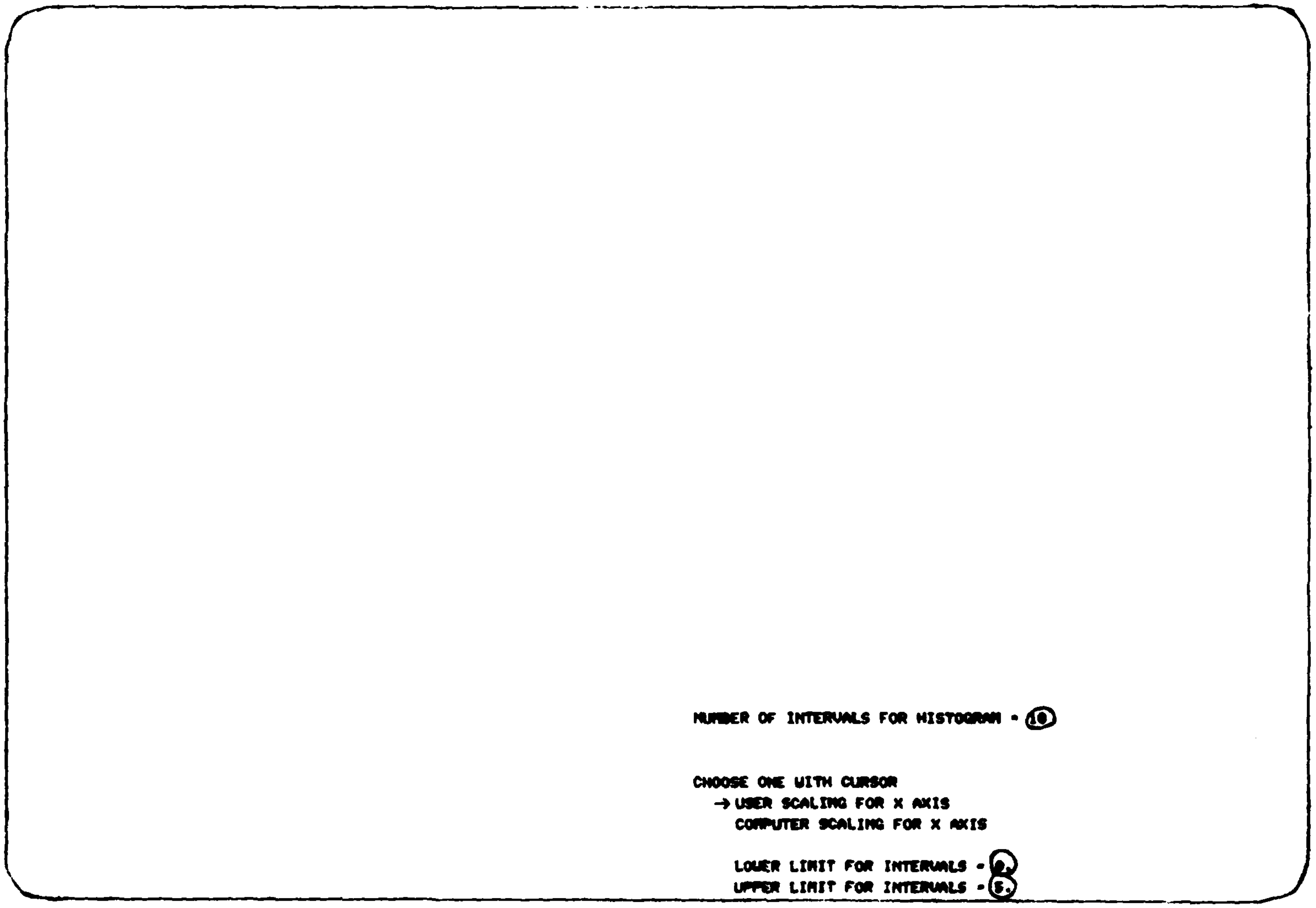


Figure 12. User scaling of histogram

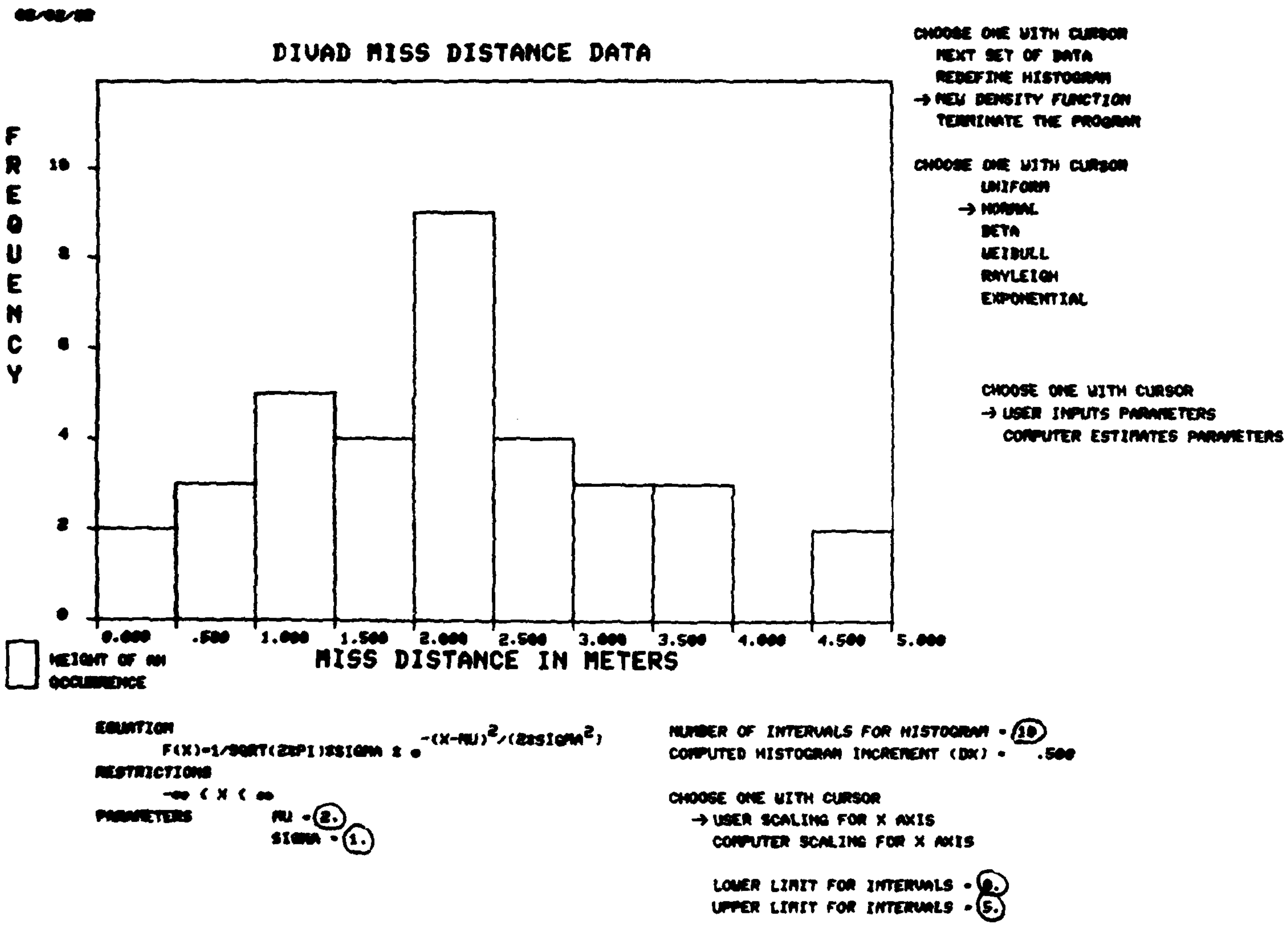


Figure 13. User inputs distribution parameters

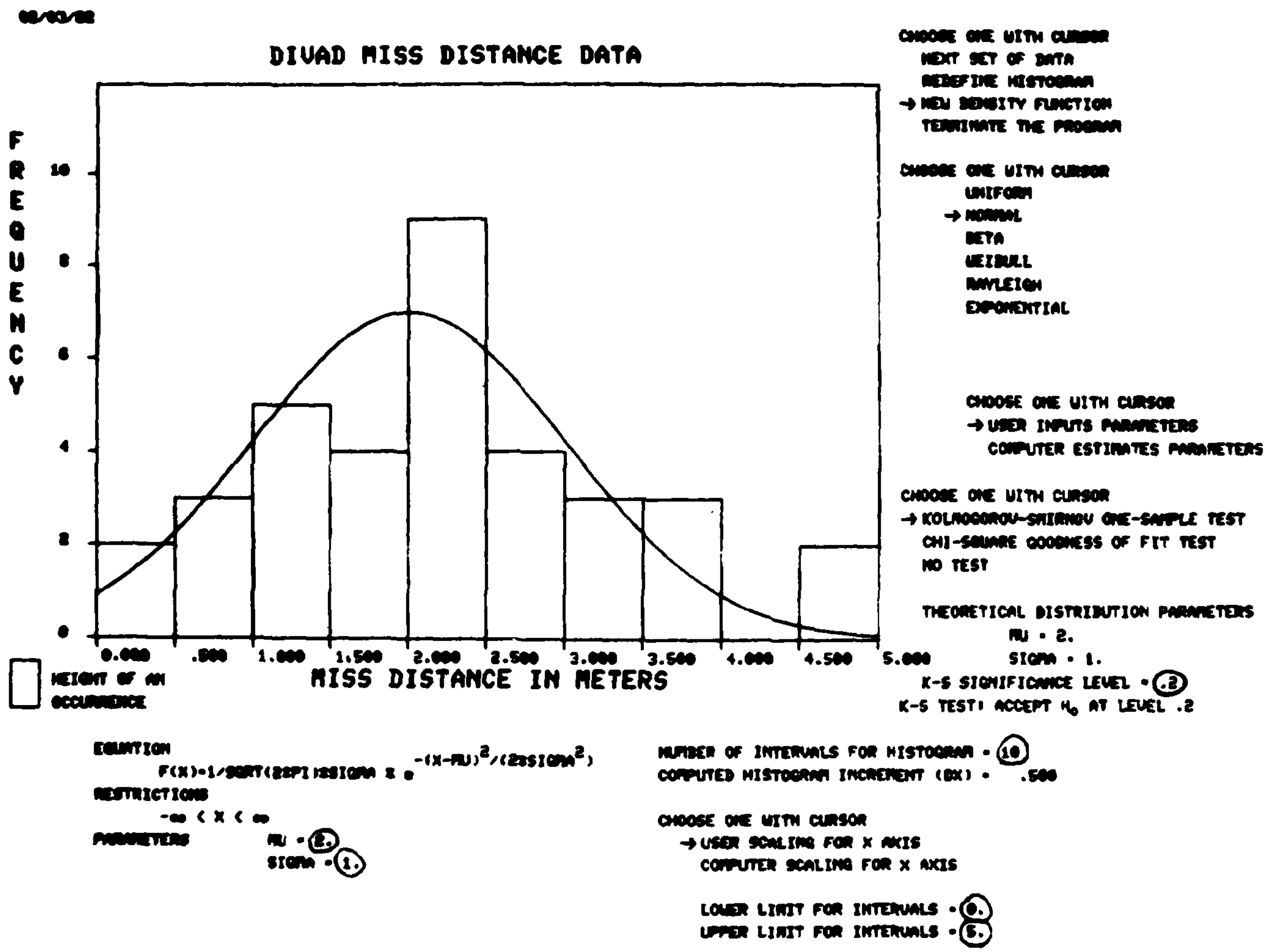


Figure 14. Inputting Kolmogorov-Smirnov significance level

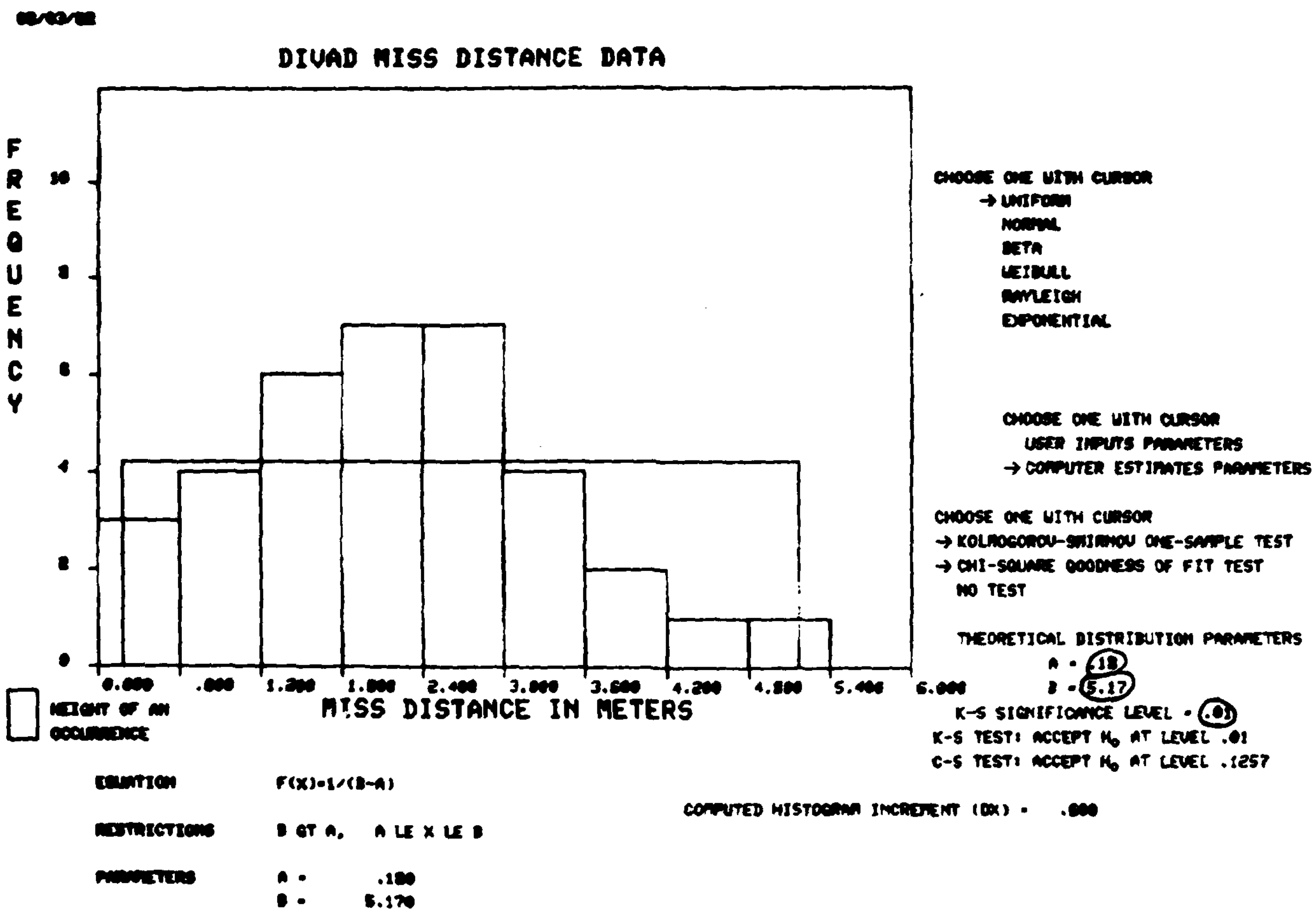


Figure 15. Testing the uniform distribution

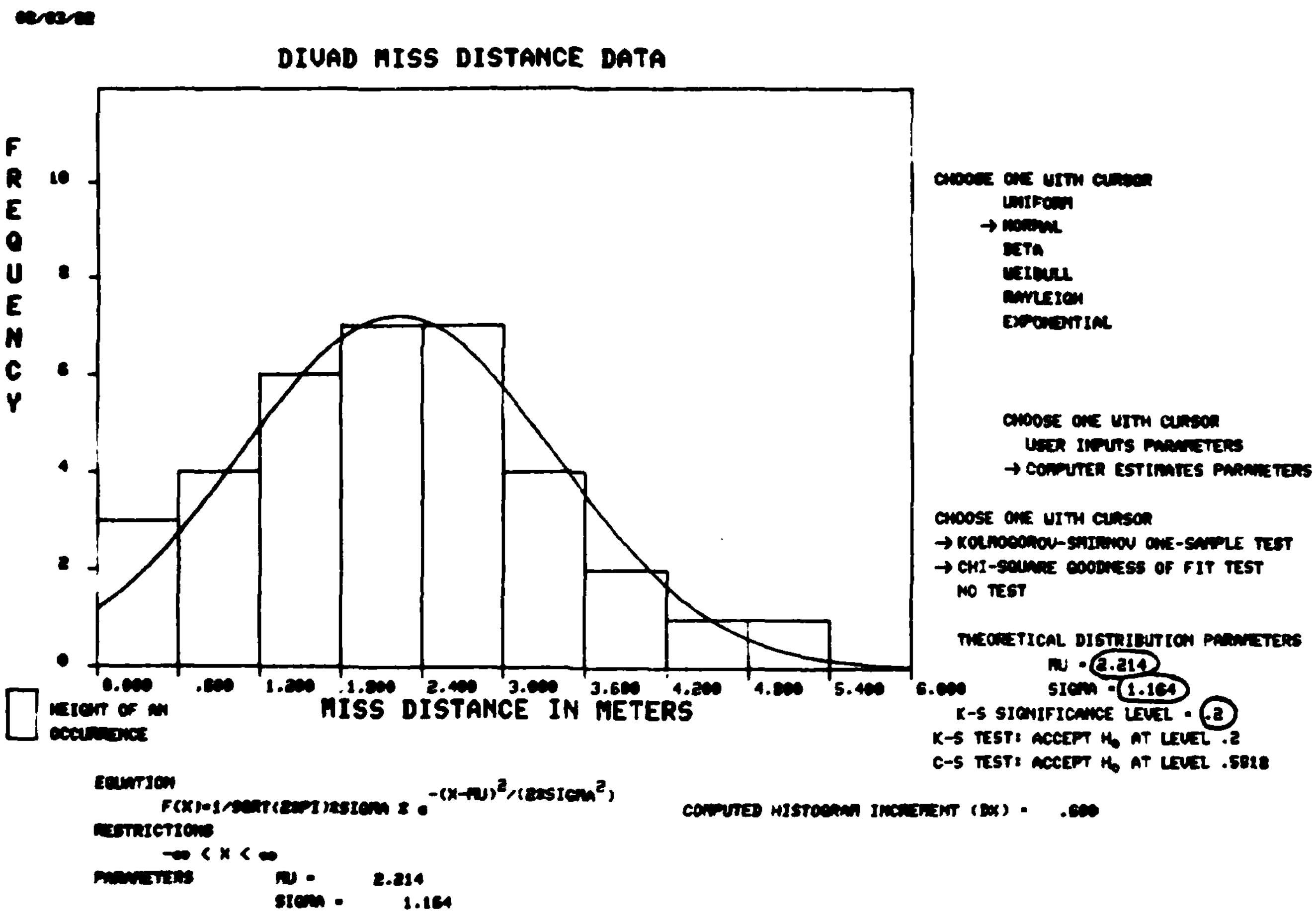


Figure 16. Testing the normal distribution

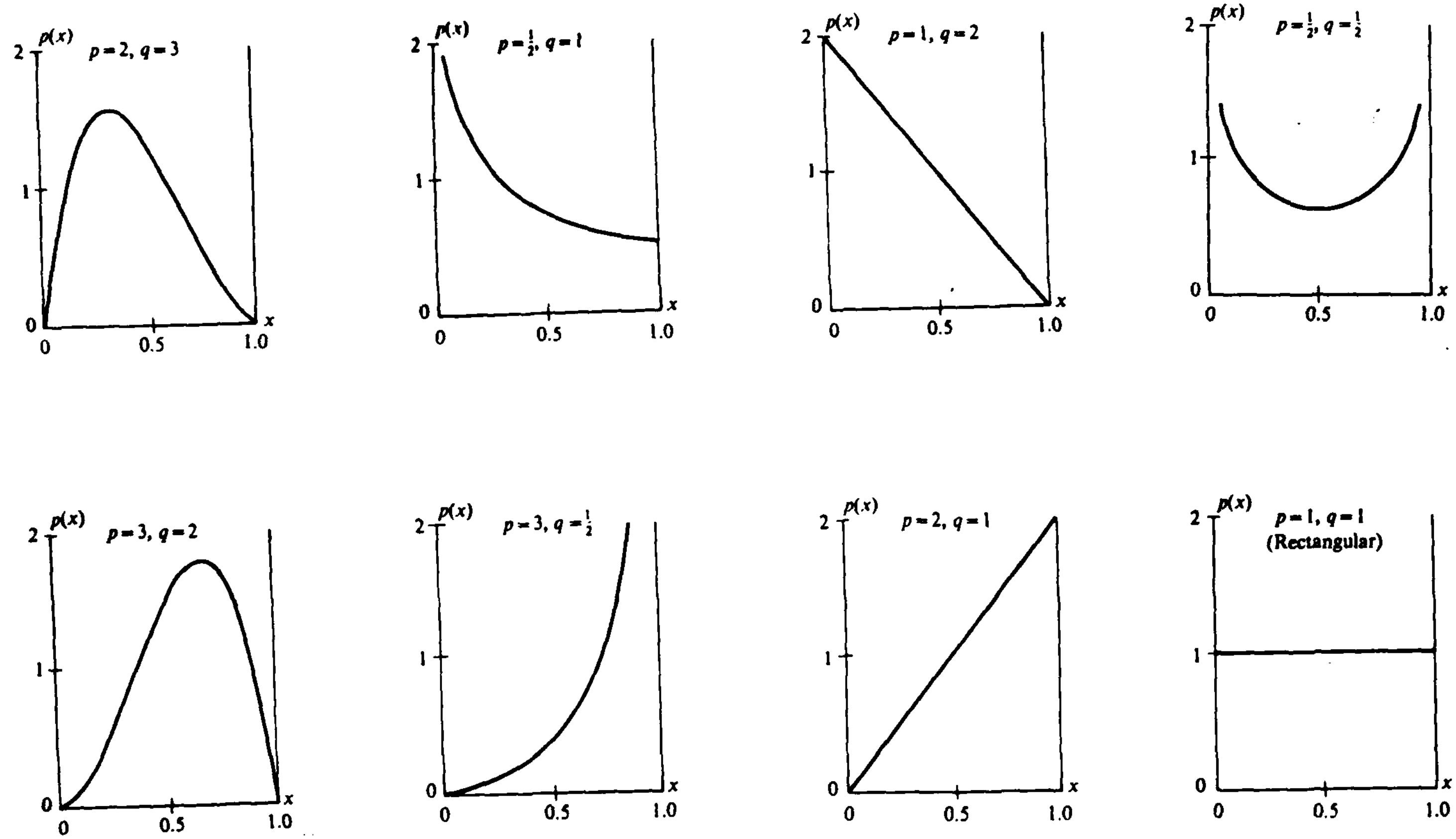


Figure 17. Beta density functions

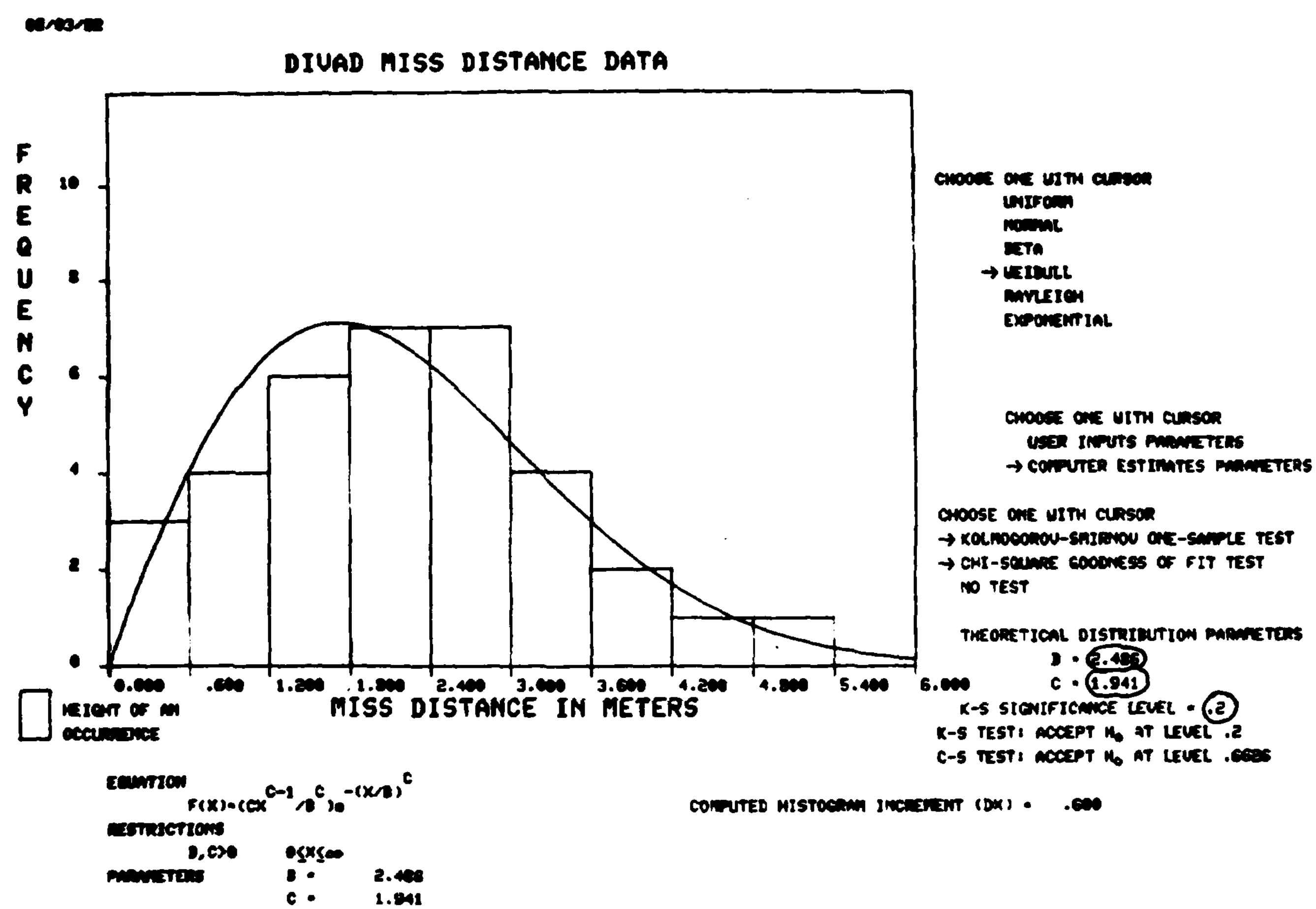


Figure 18. Testing the Weibull distribution

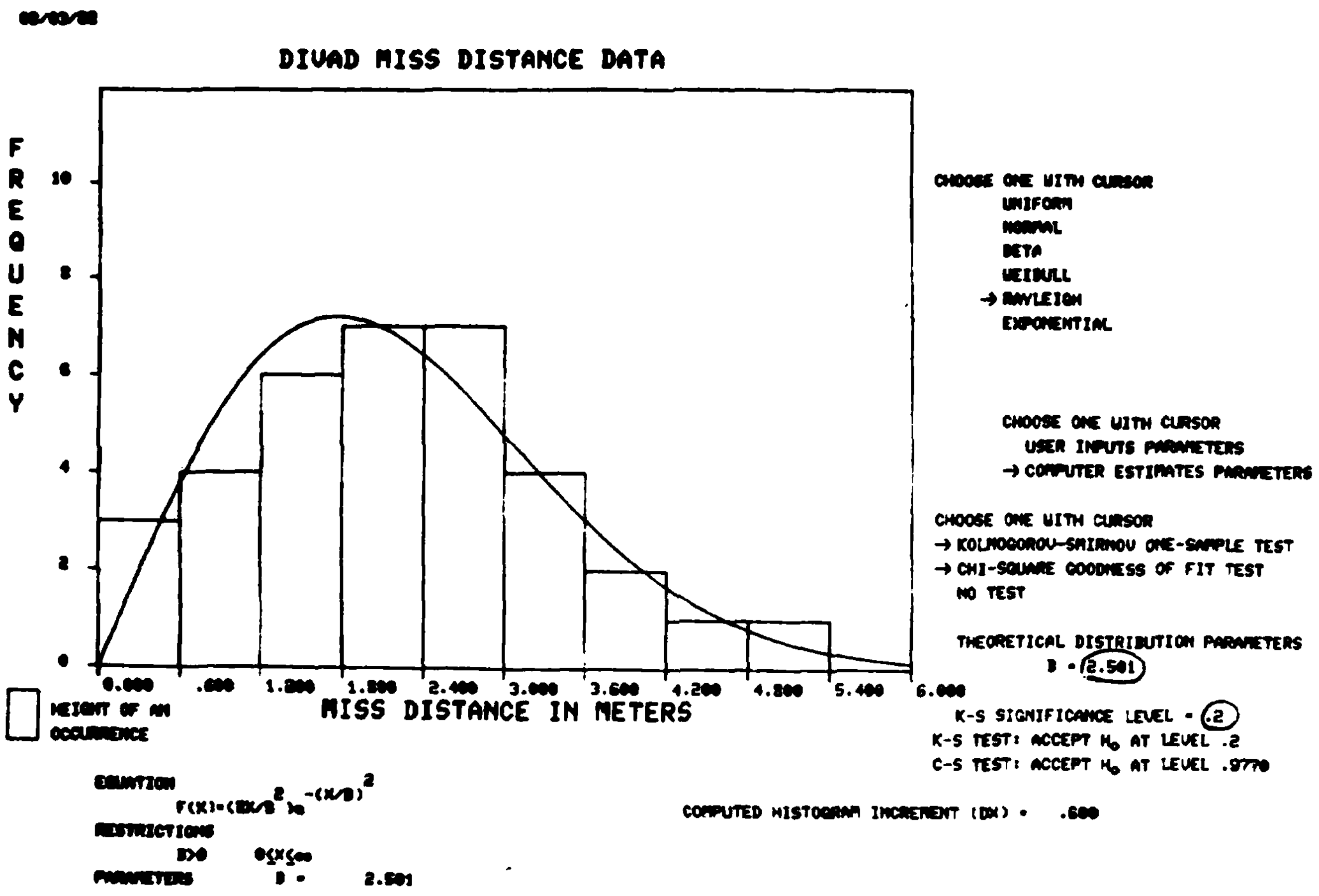


Figure 19. Testing the Rayleigh distribution

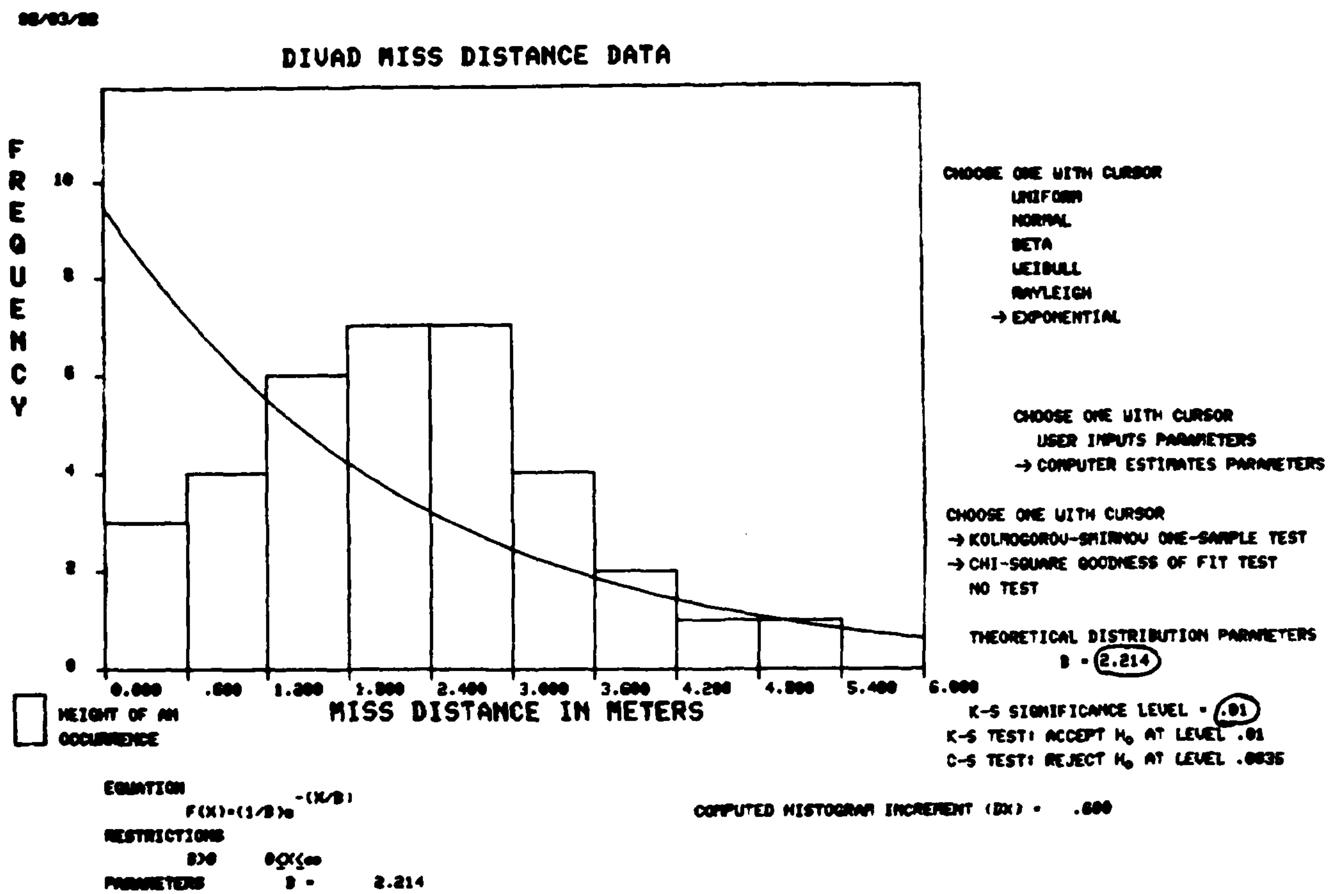


Figure 20. Testing the exponential distribution

APPENDIX A

DATA FILE CREATION PROGRAM EXAMPLE

```
CONTROL DATA INTERCOM 5.1
DATE  06/28/82
TIME  10.53.57.
                                        YOUR OWN USER NAME
PLEASE LOGIN
LOGIN,LGEMYD3300,SUP
[redacted] ENTER PASSWORD-              YOUR OWN PASSWORD
ENTER HARDWARE ID-JUX                   HARDWARE ID OF THE TERMINAL YOU ARE USING
COMMAND- ETL,100
COMMAND- SCREEN,80
COMMAND- REQUEST,TAPE5,*PF
COMMAND- ATTACH,A,PROGDATA,ID=DRDAR,CY=1,MR=1
 AT CY= 001 SN=PFSET
COMMAND- A
ENTER HORIZONTAL AXIS LABEL MISS DISTANCE IN METERS
ENTER VERTICAL AXIS LABEL FREQUENCY
ENTER DIAGRAM TITLE VULCAN MISS DISTANCE DATA
ENTER INTEGER VALUE FOR NUMBER OF DATA POINTS (1000 MAXIMUM) 32
ENTER ALL DATA POINTS AND SEPERATE EACH DATA POINT BY ONE OR MORE BLANKS
(FOR EXAMPLE: 10. .09 2.)_
          3.77 .72 1.05 1.43 .28 2.08 2.47 2.58 3.32 1.12 2.75 1.87 2.75 .57 2.19
2.41 1.27 2.3 2.29 3.53 2.19 4.7 3.3 1.75 .18 1.51 2.33 1.26 2.12 1.56 5.17 .72
ENTER: 1 TO CHANGE ANY OF THE ENTRIES
       2 TO DISPLAY THE DATA
       3 TO TERMINATE THE PROGRAM
        2
```

```
HORIZONTAL AXIS LABEL: MISS DISTANCE IN METERS

VERTICAL AXIS LABEL: FREQUENCY

DIAGRAM TITLE: VULCAN MISS DISTANCE DATA

NUMBER OF DATA POINTS:   32

DATA POINTS:
     3.770      .720     1.050     1.430      .280     2.080     2.470
     2.580     3.320     1.120     2.750     1.870     2.750      .570
     2.190     2.410     1.270     2.300     2.290     3.530     2.190
     4.700     3.300     1.750      .180     1.510     2.330     1.260
     2.120     1.560     5.170      .720

ENTER: 1 TO CHANGE ANY OF THE ENTRIES
       2 TO DISPLAY THE DATA
       3 TO TERMINATE THE PROGRAM
        1

ENTER: 1 TO CHANGE THE HORIZONTAL AXIS LABEL
       2 TO CHANGE THE VERTICAL AXIS LABEL
       3 TO CHANGE THE DIAGRAM TITLE
       4 TO CHANGE THE NUMBER OF DATA POINTS
       5 TO CHANGE THE DATA POINTS
        3
ENTER DIAGRAM TITLE DIVAD MISS DISTANCE DATA
ENTER: 1 TO CHANGE ANY OF THE ENTRIES
       2 TO DISPLAY THE DATA
       3 TO TERMINATE THE PROGRAM
        1
```

```
ENTER: 1 TO CHANGE THE HORIZONTAL AXIS LABEL
       2 TO CHANGE THE VERTICAL AXIS LABEL
       3 TO CHANGE THE DIAGRAM TITLE
       4 TO CHANGE THE NUMBER OF DATA POINTS
       5 TO CHANGE THE DATA POINTS
         4
ENTER INTEGER VALUE FOR NUMBER OF DATA POINTS (1000 MAXIMUM)_35
ENTER: 1 TO CHANGE ANY OF THE ENTRIES
       2 TO DISPLAY THE DATA
       3 TO TERMINATE THE PROGRAM
         1

ENTER: 1 TO CHANGE THE HORIZONTAL AXIS LABEL
       2 TO CHANGE THE VERTICAL AXIS LABEL
       3 TO CHANGE THE DIAGRAM TITLE
       4 TO CHANGE THE NUMBER OF DATA POINTS
       5 TO CHANGE THE DATA POINTS
         5
ENTER INTEGER VALUE FOR THE NUMBER OF DATA POINTS YOU WANT TO CHANGE_5
ENTER LIST OF DATA POINTS YOU WANT TO CHANGE, FOLLOWED BY THE CHANGES, SEPERATED
BY BLANKS (I.E., AN ENTRY OF 3 8 10 .10 2. 3.5 WILL CHANGE THE THIRD, EIGHTH,
AND TENTH DATA POINTS TO .10, 2., AND 3.5 RESPECTIVELY_)
         21 27 33 34 35 3.19 2.43 2.25 3.79 2.81
ENTER: 1 TO CHANGE ANY OF THE ENTRIES
       2 TO DISPLAY THE DATA
       3 TO TERMINATE THE PROGRAM
         2
```

```
HORIZONTAL AXIS LABEL: MISS DISTANCE IN METERS

VERTICAL AXIS LABEL: FREQUENCY

DIAGRAM TITLE: DIVAD MISS DISTANCE DATA

NUMBER OF DATA POINTS:   35

DATA POINTS:
    3.770      .720     1.050     1.430      .280     2.080     2.470
    2.580     3.320     1.120     2.750     1.870     2.750      .570
    2.190     2.410     1.270     2.300     2.290     3.530     3.190
    4.700     3.300     1.750      .180     1.510     2.430     1.260
    2.120     1.560     5.170      .720     2.250     3.790     2.810


ENTER: 1 TO CHANGE ANY OF THE ENTRIES
       2 TO DISPLAY THE DATA
       3 TO TERMINATE THE PROGRAM
        3
      END  DATA
     031300 MAXIMUM EXECUTION FL.                           OWNER OR CREATOR OF FILE
      1.067 CP SECONDS EXECUTION TIME.
COMMAND- CATALOG,TAPE5,DIVADDATA,ID=MAYERNIK,CY=1
 INITIAL CATALOG
 RP = 050 DAYS
 CT ID= MAYERNIK PFN=DIVADDATA
 CT CY= 001 SN=PFSET    0000000064 WORDS.
COMMAND- LOGOUT
CPA      2.948 SEC.         2.947 ADJ.
SYS TIME                    4.501
EST. COST AT $250/HR. - $      .31
CONNECT TIME     0 HRS.    17 MIN.
 06/28/82  LOGGED OUT AT 11.11.20.
      <
```

APPENDIX B

MULTIPLE DATA SETS PER DATA FILE

```
CONTROL DATA INTERCOM 5.1
DATE  07/12/82
TIME  12.50.07.
                                        YOUR OWN USER NAME
PLEASE LOGIN
LOGIN,LGEMYD3300,SUP
[redacted] ENTER PASSWORD-              YOUR OWN PASSWORD
ENTER HARDWARE ID-JUX                   HARDWARE ID OF THE TERMINAL YOU ARE USING
COMMAND- ETL,100
COMMAND- SCREEN,80
COMMAND- REQUEST,TAPE5,*PE
COMMAND- ATTACH,A,PROGDATA,ID=DRDAR,CY=1,MR=1
 AT CY= 001 SN=PFSET
COMMAND- A
ENTER HORIZONTAL AXIS LABEL COST IN DOLLARS
ENTER VERTICAL AXIS LABEL FREQUENCY
ENTER DIAGRAM TITLE OLD TYPE CONTAINER
ENTER INTEGER VALUE FOR NUMBER OF DATA POINTS (1000 MAXIMUM) 4
ENTER ALL DATA POINTS AND SEPERATE EACH DATA POINT BY ONE OR MORE BLANKS
(FOR EXAMPLE: 10. .09 2.)
        12.44 17.11 14.27 12.14
ENTER: 1 TO CHANGE ANY OF THE ENTRIES
       2 TO DISPLAY THE DATA
       3 TO TERMINATE THE PROGRAM
        3
      END  DATA
     031300 MAXIMUM EXECUTION FL.
       .487 CP SECONDS EXECUTION TIME.
COMMAND- A
```

```
ENTER HORIZONTAL AXIS LABEL_COST IN DOLLARS
ENTER VERTICAL AXIS LABEL_FREQUENCY
ENTER DIAGRAM TITLE_NEW TYPE CONTAINER
ENTER INTEGER VALUE FOR NUMBER OF DATA POINTS (1000 MAXIMUM)_4
ENTER ALL DATA POINTS AND SEPERATE EACH DATA POINT BY ONE OR MORE BLANKS
(FOR EXAMPLE: 10. .09 2.)_
     13.52 18.08 15.36 13.24
ENTER: 1 TO CHANGE ANY OF THE ENTRIES
       2 TO DISPLAY THE DATA
       3 TO TERMINATE THE PROGRAM
        3
      END  DATA
     031300 MAXIMUM EXECUTION FL.                    OWNER OR CREATOR OF FILE
       .449 CP SECONDS EXECUTION TIME.
COMMAND- CATALOG,TAPE5,PACKAGINGDATA,ID=MAYERNIK,CY=1
 INITIAL CATALOG
 RP = 050 DAYS
 CT ID= MAYERNIK PFN=PACKAGINGDATA
 CT CY= 001 SN=PFSET    0000000064 WORDS.
COMMAND- LOGOUT
CPA       3.689 SEC.        3.688 ADJ.
CPB        .000 SEC.         .000 ADJ.
SYS TIME                    6.554
EST. COST AT $250/HR. - $     .45
CONNECT TIME     0 HRS.    9 MIN.
 07/12/82  LOGGED OUT AT 12.59.10.
      <
```

DISTRIBUTION LIST

Commander
U.S. Army Armament Research and
Development Command
ATTN: DRDAR-TSS (5)
DRDAR-RAA
DRDAR-LC
DRDAR-SC
DRDAR-MS
DRDAR-MSM
DRDAR-TS
DRDAR-QA
DRDAR-PS
DRDAR-PM
DRCPM-CW
DRCPM-ADG
DRCPM-SMK
SARPM-PBM
DRCPM-NUC
DRDAR-AS
DRDAR-NC
DRDAR-RAC (16)
DRCPM-TMA
DRXHE-AR
Dover, NJ 07801

Administrator
Defense Technical Information Center
ATTN: Accessions Division
Cameron Station
Alexandria, VA 22314 (12)

Director
U.S. Army Materiel Systems Analysis Activity
ATTN: DRXSY-MP
Aberdeen Proving Ground, MD 21005

Commander
U.S. Army Armament Research and
Development Command
Weapons Systems Concepts Team
ATTN: DRDAR-ACW
APG, Edgewood Area, MD 21010

Commander/Director
Chemical Systems Laboratory
U.S. Army Armament Research and
Development Command
ATTN: DRDAR-CLJ-L
APG, Edgewood Area, MD 21010

Director
Ballistics Research Laboratory
U.S. Army Armament Research and
Development Command
ATTN: DRDAR-TSB-S
Aberdeen Proving Ground, MD 21005

Chief
Benet Weapons Laboratory, LCWSL
U.S. Army Armament Research and
Development Command
ATTN: DRDAR-LCB-TL
Watervliet, NY 12189

Commander
U.S. Army Armament Materiel
Readiness Command
ATTN: DRSAR-LEP-L
Rock Island, IL 61299

Director
U.S. Army TRADOC Systems
Analysis Activity
ATTN: ATAA-SL
White Sand Missile Range, NM 88002

Commander
U.S. Army DARCOM
ATTN: DRCCP-E
Alexandria, VA 22333

Comptroller of the Army
ATTN: DACA-CAZ-A
Pentagon Building
Washington, D.C. 20301